September 2004, Volume 8, Number 3
pp. 122-143

TESTING L2 VOCABULARY RECOGNITION AND RECALL USING PICTORIAL AND WRITTEN TEST ITEMS

Linda Jones
University of Arkansas

ABSTRACT

This article describes two studies that examined the effects of pictorial and written annotations on second language (L2) vocabulary learning from a multimedia environment. In both studies, students were randomly assigned to one of four aural multimedia groups: a control group that received no annotations, and three treatment groups that provided written, pictorial, or both written and pictorial annotations while listening. In the first study, students in the three treatment groups recognized English translations or pictorial representations of French keywords better than the control group that received no annotations during listening. In the second study, students produced English translations of French keywords best when the mode of testing matched the treatment mode. These results add to the growing body of literature on the beneficial effects of annotations on L2 vocabulary recognition and recall.

INTRODUCTION

In recent years, media-based listening comprehension activities have evolved from a purely audio-only approach to one that is more holistic and multi-sensory. No longer are materials focused on nonsensical sentence structures. Instead, students now experience lengthier, authentic audio passages embedded in video, interactive CD-ROMs, or Web sites. Numerous computer-based and online packages have been developed by researchers, faculty, and publishing companies (Amon, Muyskens, & Omaggio Hadley, 2000; Chun & Plass, 1997; Larson & Bush, 1992; Otto & Pusack, 1992; Sabo, Restrepo, & Jones, 2000; University of Texas, 2001, 2004) to assist students' L2 aural and written development. *Français Interactif* (University of Texas, 2001, 2004) is one of the more innovative on-line French language learning programs produced in recent years. It provides multiple levels of semester-long instruction with a mixture of multi-sensory materials such as aural, pictorial, video, and written presentations that help students to learn the target language. Interactive self-checking exercises provide them with an opportunity to examine their output in terms of recall of the target language material. *Cyberbuch* (Chun & Plass, 1997), another innovative program on CD-ROM, provides advance organizer videos and annotated information to support students' reading of a German text. This highly focused program promotes interaction with multi-sensory annotated information in the form of text translations, images, short video clips, and audio to facilitate students' understanding of keywords and the literary texts.

As the development of L2 multimedia packages such as *Français Interactif* and *Cyberbuch* increases, researchers strive to better understand how the attributes of multimedia can enhance listening and reading comprehension and vocabulary acquisition (Brett, 1995, 1997; Hoven, 1999; Jones & Plass, 2002; Lynch 1998; Salaberry, 2001). For example, Chun and Plass (1996) examined the influence of written and pictorial annotations on students' vocabulary learning from a written text whereas Jones and Plass (2002) examined their influence on vocabulary learning from an aural text. Other researchers, for example Doughty (1991) and Pica, Doughty, and Young (1986) studied the effects of student control over the L2 multimedia environment, while researchers such as Herron (1994) and Herron, York, Cole, and Linden (1998) have closely examined the influence of advance organizers on L2 learning. Despite these advances, many questions remain concerning the effects of multimedia components on students' L2 development.

This article expands upon previous research (Jones, 2003; Jones & Plass, 2002) by describing two studies that used three separate vocabulary tests (written recognition, pictorial recognition, and written recall) to examine how pictorial and written test items might demonstrate how well written and pictorial annotations assist in the learning of new vocabulary when students listen to an aural L2 passage in a multimedia environment. The present article examines this topic. It begins with a review of the role of written and pictorial annotations in L2 reading and listening comprehension, of the process of incidental vocabulary learning, and of the different methods and tests used to examine students' vocabulary recognition and recall. The article concludes with a discussion of the findings, the implications and limitations of this study, and suggestions for future research.

VOCABULARY LEARNING FROM ANNOTATIONS

Information is cognitively processed through visual or verbal channels (Mayer, 1997, 2001, 2002; Paivio, 1986). A dual processing strategy assumes that individuals develop mental pictorial representations of graphic input and mental verbal representations of linguistic input. The presence of both pictorial and verbal cues can facilitate learning, in particular when the corresponding visual and verbal representations are contiguously present in working memory (Mayer, 1997, 2001, 2002; Wittrock, 1989). Mayer's generative theory of multimedia learning (1997, 2001) states that in order to meaningfully comprehend a text in a multimedia format, learners must select relevant pictorial and/or linguistic information from it, organize the input into coherent visual and verbal mental representations, and then integrate the latter by constructing referential connections between the two.

Researchers have long been interested in examining the effects of pictorial and verbal cues on L2 vocabulary learning, and have found that processing supportive information such as pictures or translations enhances language learning. With regard to high-imagery concrete vocabulary learning, Kellogg and Howe (1971) found that foreign words associated with images or actual objects are learned more easily than those without such additional information. Terrell (1986) found that combining an unknown L2 word with a visual representation bypasses a direct translation and facilitates vocabulary learning. Underwood (1989) suggested that we "remember images better than words, hence we remember words better if they are strongly associated with images" (p. 19). Other research suggests that foreign words associated with aural or written translations and images are learned more easily than are those accompanied by pictures or text alone (Baltova, 1999; Guillory, 1998; Jones & Plass, 2002; Oxford & Crookall, 1990; Plass, Chun, Mayer, & Leutner, 1998). For example, Oxford and Crookall (1990) suggest that the combination of pictures and text accesses more parts of the brain, thereby leading to greater depth of processing than when text is processed alone. Baltova (1999) examined the effects of viewing a French video with either French audio and French subtitles (bimodal format), or English audio and French subtitles (reversed format) on students' vocabulary learning. She found that students learned significantly more vocabulary when they viewed the audio-visual material with both French subtitles and French audio present than in the reversed format where they viewed the video with English audio and French subtitles.

Incidental vocabulary learning is the process of acquiring vocabulary while reading or listening for comprehension rather than focusing solely on memorizing lists of words (Hulstijn, 1989; Hulstijn, Hollander, & Greidanus, 1996; Yoshii & Flaitz, 2002). Any incidental vocabulary learning that occurs in a multimedia environment may depend upon the type of annotations processed, and the depth of experience with them. For example, Hulstijn, Hollander, and Greidanus (1996) examined how the presence of glosses for targeted words, or dictionary lookup of words in a written text, might affect incidental vocabulary learning. They found that incidental learning of words frequently occurring in the text was more likely when learners were provided with access to word meanings through marginal glosses or dictionaries than when no helpful information was made available to them. Hulstijn (1992) determined that deep elaboration of the meaning of an unknown word also led to incidental vocabulary learning. Jones (2003) found that students believed that pictures demanded deeper processing than did verbal

translations because they had to "figure out" the meaning which they did not have to do if they saw the translation immediately. Quantitative results confirmed her beliefs: Students who accessed pictorial annotations demonstrated greater incidental vocabulary learning than those who did not access this annotation type. Other researchers found that if the context of a written or spoken passage was not clear from the onset, deeper processing might fail to support incidental vocabulary learning, and students who do not have access to annotated information would run the risk of learning words incorrectly (Chun & Plass, 1996; Hulstijn, 1992; Jones, 2003; Jones & Plass, 2002).

Recognition and recall tests are often used to examine students' vocabulary knowledge. However, test and measurement studies indicate that these two forms of testing are quite different and demand separate processing strategies (Cariana & Lee, 2001; Jonassen & Tessmer, 1996). For example, recognition tests usually involve multiple choice activities whereby learners select or guess the correct response from the alternatives given. Such tests may strengthen any existing memory traces (McDaniel & Mason, 1985). Recall, on the other hand, demands the production of responses from memory. It is more difficult than recognition because learners must search for the correct response within their mental representation of the newly experienced information (Cariana & Lee, 2001; Glover, 1989; McDaniel & Mason, 1985).

Several studies have investigated the use of pictorial and written annotations in L2 multimedia reading and listening comprehension using different testing formats (Chun & Plass, 1996; Jones & Plass, 2002; Plass et al., 1998). Plass et al. found that when students accessed both pictorial and written annotations as they read a multimedia-based German text, they scored higher on a written vocabulary production test than when only one annotation type was accessed. The combination of both annotation types allowed for more than one retrieval route to the information in long term memory. These researchers also found that written annotations had a stronger impact on vocabulary production than did pictorial annotations. Jones and Plass (2002) reported that those students who accessed both pictorial and written annotations as they listened to a multimedia-based aural French text performed better on a written vocabulary recognition test than those who accessed single annotations, or no annotations at all. However, unlike subjects in the Plass et al. study, those who accessed pictorial annotations alone or combined with written annotations outperformed those who did not access pictorial annotations on a written vocabulary recognition test. Chun and Plass (1996) further examined the effects of multimedia annotations on L2 vocabulary learning from a reading passage using a written production and a recognition test with a balance of pictorial and written test items that paralleled the modality in which the information was presented. They, too, found that students performed best on both types of tests when both pictorial and written annotations were viewed than when single or no annotations were accessed during reading. They also observed that when the method of testing more closely paralleled the way in which information was presented, student performance improved considerably, resulting in 77% of correct responses on immediate and delayed vocabulary tests, a percentage much higher than the 23%-55% typically expected in select-definition tests (Knight, 1994).

All of the above studies suggest incidental vocabulary learning can be increased if learners are given opportunities to look up word meanings, visually or verbally, while listening or reading. However, none of these studies specifically examined students' incidental vocabulary learning from a listening comprehension activity using vocabulary tests that complemented or ignored the annotation type accessed.

The following two studies, therefore, investigated how pictorial and/or written annotations affect students' performance on incidental vocabulary learning tests that required them to either recognize or recall vocabulary incidentally learned from an aural text, using pictorial or written test items. Three hypotheses are thus proposed that coincide with the three dependent measures used in this study and introduced in the next section:

1. Students with access to pictorial and written annotations during a L2 listening comprehension activity will recognize more written translations of keywords on a written vocabulary recognition posttest than those with access to one type of annotation, or no annotations at all. In addition, students who access written annotations will outperform those without such access on a written vocabulary recognition posttest.

2. Students with access to pictorial and written annotations during a L2 listening comprehension activity will recognize more pictorial representations of keywords on a pictorial vocabulary recognition posttest than those with access to one type of annotation, or no annotations. In addition, students who access pictorial annotations will outperform those without such access on a pictorial vocabulary recognition posttest.

3. Students with access to pictorial and written annotations during a L2 listening comprehension activity will recall more keyword translations on a written vocabulary production posttest than students with access to only one type of annotation or no annotations. In addition, students who access written annotations will outperform those without access to such annotations on a written vocabulary production posttest.

THE PRESENT STUDIES

Method

Participants, Study 1

Eighty second-semester English-speaking beginning students of French, enrolled at the University of Arkansas in the fall of 2001, voluntarily participated in the study during their regular class time. The students completed a 25-item vocabulary recognition pre-test to determine their prior knowledge of the vocabulary in this study. All students demonstrated low prior knowledge of the vocabulary with an average score of 4 out of a maximum score of 25 (M = 4.04, SD = 3.60; Table 1). A Tukey HSD (honestly significant difference) multiple comparison test showed no significant differences among the four groups.

Table 1. Vocabulary Pretest Results Based on Random Assignments to Four Treatments, Study 1

Groups	*N*	*M*	*SD*
Control	20	4.15	3.50
Pictorial Annotations	20	4.15	2.94
Written Annotations	20	4.75	4.38
Pictorial and Written Annotations	20	3.10	3.51

Participants, Study 2

Sixty seven second-semester English-speaking beginning students of French, enrolled at the University of Arkansas in the spring of 2002, voluntarily participated in the study during their regular class time. They completed a 25-item pre-treatment vocabulary recognition test based on the words used in this study and demonstrated low prior knowledge of the vocabulary with an average score of 1.5 out of 25 (M = 1.57, SD = 1.23; Table 2). A Tukey HSD multiple comparison test showed no significant differences among the four groups.

Table 2. Vocabulary pretest results based on random assignments to four treatments, Study 2

Groups	*N*	*M*	*SD*
Control	16	1.69	1.45
Pictorial Annotations	17	1.24	1.03
Written Annotations	18	1.67	1.41
Pictorial and Written Annotations	16	1.69	1.01

Materials and Apparatus for Studies 1 and 2

Four aural multimedia treatments, developed using Adobe Premiere 4.2 (Adobe, 1994) and Authorware 4.0 (Macromedia, 1997), were presented to students using a 24-station Macintosh computer lab, arranged so that the students could view only their own computer screens.

All groups first saw an opening screen that instructed them how to use the program and provided an advance organizer in the form of a brief written paragraph that placed the aural passage about an important event in its historical context (Figure 1). This screen provided additional instructions to assist students with the annotations available in their respective treatments.

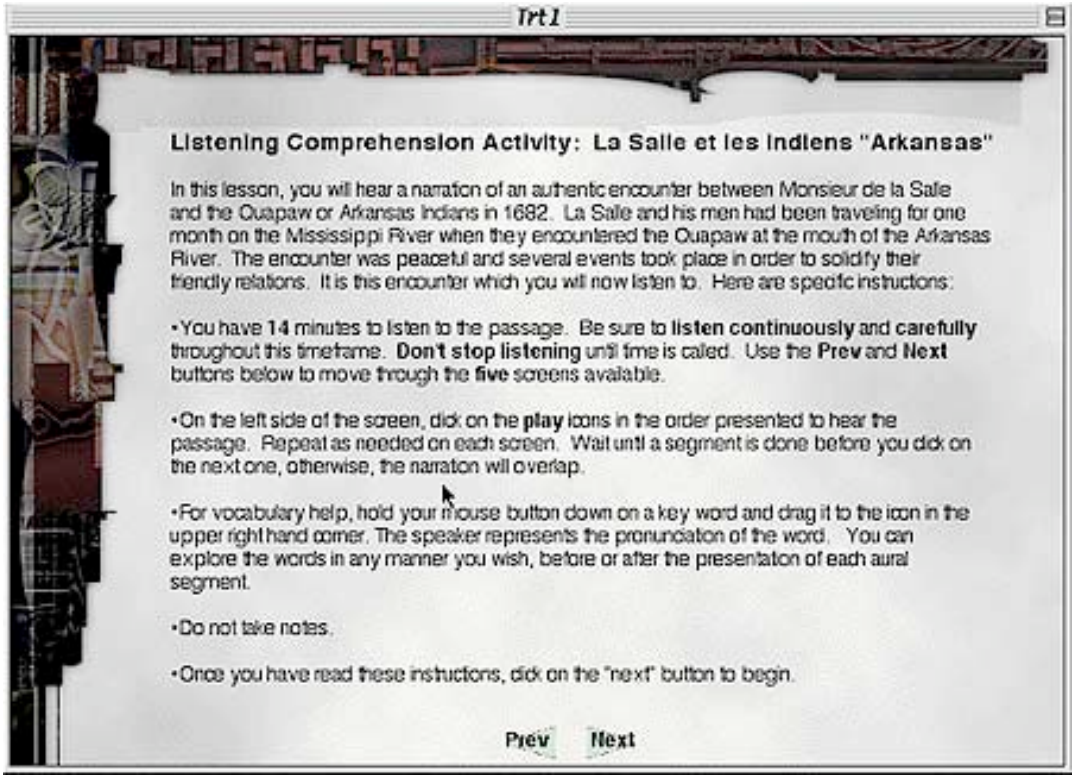


Figure 1. Example of the opening screen which provides instructions and advance organizer information, based on the treatment, prior to listening to the aural passage

The opening screen was followed by five separate listening comprehension screens tailored to each treatment. Within each screen, students could click on audio buttons to listen to a 2 minute and 20 second aural passage (Buzhardt & Hawthorne, 1993; see Appendix A). Twenty-seven French keywords, including nouns, verbs, adjectives, and adverbial phrases, were chosen by two experienced French professors for their importance to understanding the story. They were placed in order of appearance on the left side of each screen. To hear their pronunciation, listeners dragged the keywords to a speaker icon in the upper right section of the screen.

In the control group, students could only listen to the pronunciation of French keywords (Figure 2).

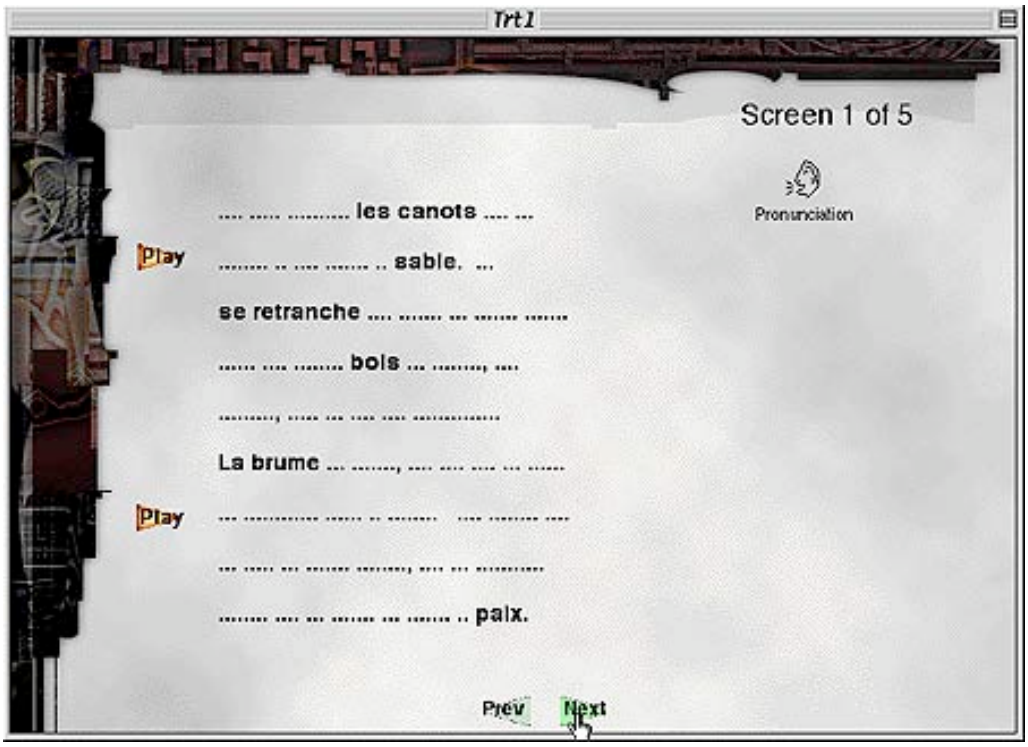


Figure 2. Example of control group treatment (no annotations available)

In the pictorial annotations group, students could drag the keywords to a camera icon to view their pictorial representations (Figure 3).

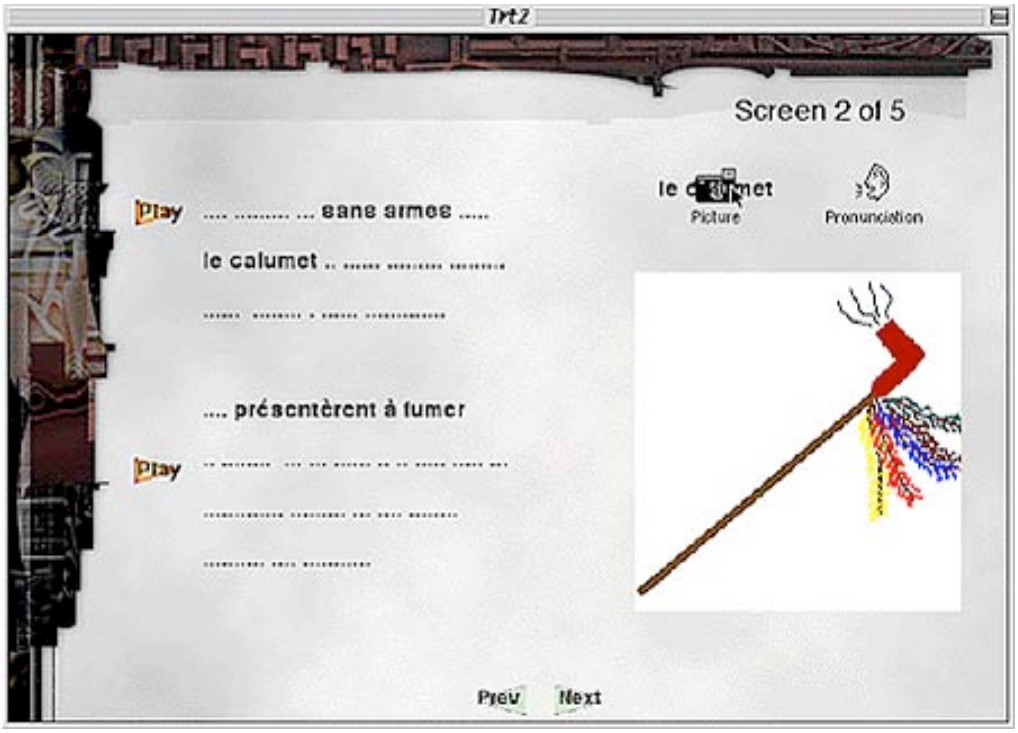


Figure 3. Example of treatment for the pictorial annotations group (only pictorial annotations available)

In the written annotations group, students could drag the keywords to a book icon to view their English translations (Figure 4).

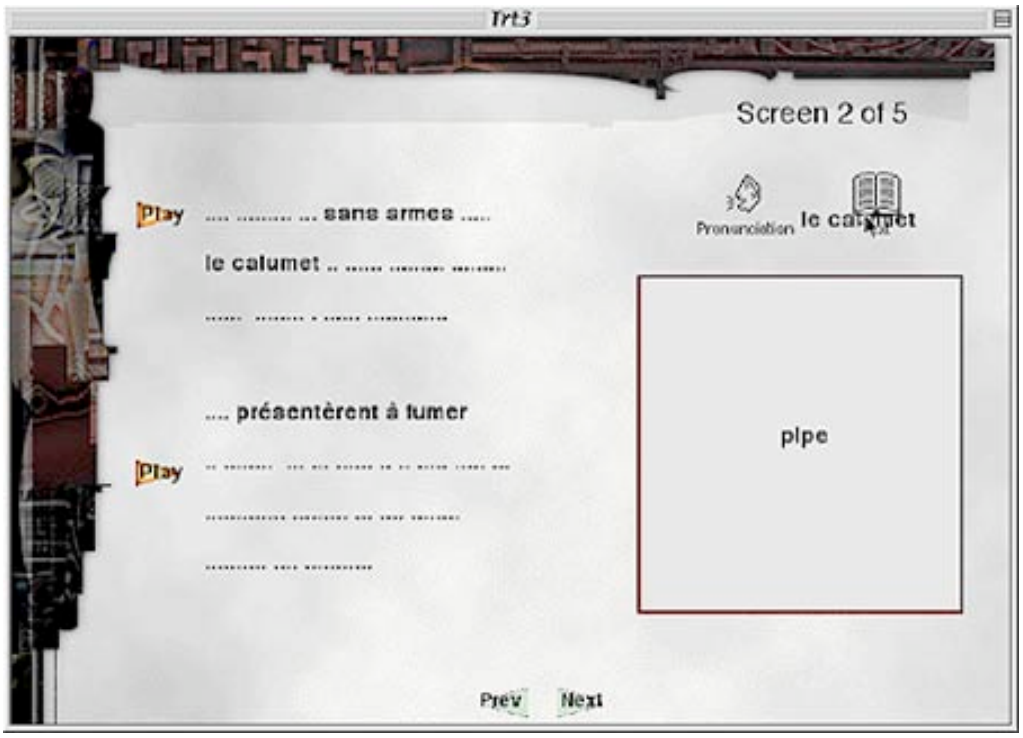


Figure 4. Example of treatment for the written annotations group (only written annotations available)

In the pictorial and written annotations group, students could drag the keywords to the camera and/or book icon to view the picture and/or an English translation (Figure 5).

Figure 5. Example of treatment for the written and pictorial annotations group

Students could select any annotation available in their treatment at any time before, during, or after each aural segment. A login script tracked the number of annotations accessed and the amount of time spent on each.

The written English translations of the French keywords were presented in a bold, 14-point Helvetica font. The color drawings and photos were pictorial representations of the same French keywords (Appendix B). The pictures were selected based on a pilot study conducted in the summer of 1999 and were used in two subsequent studies (Jones, 2003; Jones & Plass, 2002). While all written annotations provided exact English translations of the French keywords, pictorial annotations may not have precisely represented the meaning of some low-imagery French keywords such as *étouffé* (smother).

Dependent Measures and Scoring

Not all posttests were conducted in both studies since they were originally meant to study unrelated issues. However, the results of these two studies warranted a combined analysis, albeit not a full one.

In Study 1, two dependent variables examined the effects of the four treatments on students' L2 incidental vocabulary learning. The multiple choice written vocabulary recognition and pictorial recognition tests were administered immediately after the treatment and again three weeks later (Appendix C). They consisted of 25 of the 27 French keywords described above. The maximum score for each test was 25 with each correct response worth one point. The written vocabulary recognition pre- and posttests were identical. In these multiple choice tests, students had six English translations to select from for each test item. The pictorial vocabulary recognition posttest provided five pictorial representations to select from for each test item.

In Study 2, a written vocabulary production posttest (Appendix D) was used to examine the effects of the four treatments on students' vocabulary learning both immediately after the treatment and 3 weeks later. It consisted of 25 of the 27 French keywords used in each treatment, and required students to provide their English translations from memory. The maximum score for this test was 25, with each correct response worth one point.

Procedures for Studies 1 and 2

A pre/posttest control group design was used to observe the effects of the absence or presence of written and pictorial annotations on students' L2 vocabulary learning from the aural passage. All activities took place during three separate class periods of a normally scheduled French class. During the first class period, students had 8 minutes to complete the written vocabulary recognition pretest. Two days later, each participant was randomly assigned to one of four treatments: (1) no annotations (2) pictorial annotations, (3) written annotations, and (4) pictorial and written annotations. Students were given 14 minutes to listen to the passage and to access the annotations. Students in Study 1 then had 8 minutes each to complete the written and the pictorial recognition posttests. Students in Study 2 had 8 minutes to complete the immediate written vocabulary production posttest. Three weeks later, without any additional experience with the aural passage and without prior warning, students in both studies completed the delayed vocabulary tests that were identical to the tests given immediately after treatment.

RESULTS

Immediate Written Vocabulary Recognition Posttest, Study 1

A 2 x 2 ANOVA was computed with the number of correct answers on the immediate written vocabulary recognition posttest as the dependent measure, and the presence or absence of pictorial and written annotations as the between subjects factor (Table 3).

Table 3. ANOVA for Immediate Written Vocabulary Recognition Posttest, Study 1

Factors	*F*	*MSE*	*p*	*n2*
Written Annotations	51.0 (1,76)	931.60	<0.001	0.401
Pictorial Annotations	22.91 (1,76)	418.61	<0.001	0.232
Written and Pictorial Annotations	22.41 (1,76)	409.51	<0.001	0.228

There was a significant main effect for written annotations and for pictorial annotations, and a significant interaction effect between the two. The pictorial and written annotations group, and the written annotations group performed best while the control group performed the poorest (Table 4).

Table 4. Mean Group Scores on the Immediate Written Vocabulary Recognition Posttest, Study 1

Groups	*N*	*M*	*SD*
Control	20	7.80	3.85
Pictorial Annotations	20	16.90	3.91
Written Annotations	20	19.15	4.51
Pictorial and Written Annotations	20	19.20	4.76

Tukey HSD showed that all annotation groups performed significantly better than the control group ($p<0.001$), but did not differ significantly from each other.

Delayed Written Vocabulary Recognition Posttest, Study 1

A 2 x 2 ANOVA was computed with the number of correct answers on the delayed written vocabulary recognition posttest as the dependent measure and the presence or absence of pictorial and written annotations as the between subjects factor (Table 5).

Table 5. ANOVA for Delayed Written Vocabulary Recognition Posttest, Study 1

Factors	*F*	*MSE*	*p*	*n2*
Written Annotations	12.3 (1,64)	271.82	<0.05	0.161
Pictorial Annotations	6.03 (1,64)	133.70	<0.05	0.086
Written and Pictorial Annotations	4.17 (1,64)	92.47	<0.05	0.061

There was a significant main effect for written and for pictorial annotations, and a significant interaction effect between pictorial and written annotations. Mean group scores showed that the pictorial and written annotations group performed the best, while the control group performed the poorest (Table 6).

Table 6. Mean Group Scores on the Delayed Written Vocabulary Recognition Posttest, Study 1

Groups	*N*	*M*	*SD*
Control	17	8.06	2.93
Pictorial Annotations	19	13.42	4.48
Written Annotations	18	14.33	4.51
Pictorial and Written Annotations	14	15.00	6.64

A Tukey HSD test showed that the annotation groups had significantly higher scores than the control group ($p<0.001$). There were no statistically significant differences between the treatment groups.

Immediate Pictorial Vocabulary Recognition Posttest, Study 1

A 2 x 2 ANOVA was computed with the number of correct answers on the immediate pictorial vocabulary recognition posttest as the dependent measure and the presence or absence of pictorial and written annotations as the between subjects factor (Table 7).

Table 7. ANOVA for the Immediate Pictorial Vocabulary Recognition Posttest, Study 1

Factors	*F*	*MSE*	*p*	*n2*
Written Annotations	14.12 (1,76)	270.11	<0.001	.157
Pictorial Annotations	39.22 (1,76)	750.31	<0.001	.340
Written and Pictorial Annotations	10.21 (1,76)	195.31	<0.001	.118

There was a significant main effect for pictorial and for written annotations, and a significant interaction effect between the two. The pictorial and written annotations group and the pictorial annotations group performed the best while the control group performed the poorest (Table 8).

Table 8. Mean Scores of the Four Groups on the Immediate Pictorial Vocabulary Recognition Posttest, Study 1

Groups	*N*	*M*	*SD*
Control	20	11.80	3.65
Pictorial Annotations	20	21.05	4.38
Written Annotations	20	18.60	5.15
Pictorial and Written Annotations	20	21.60	4.17

Post hoc comparisons (Tukey HSD) of the posttest scores showed that all annotation groups had significantly higher scores than the control group ($p<0.001$), but that there were no statistically significant differences between the annotation groups.

Delayed Pictorial Vocabulary Recognition Posttest, Study 1

A 2 x 2 ANOVA was computed with the number of correct answers on the delayed pictorial vocabulary recognition posttest as the dependent measure and the presence or absence of pictorial and written annotations as the between subjects factor (Table 9).

Table 9. ANOVA for the Delayed Pictorial Vocabulary Recognition Posttest, Study 1

Factors	*F*	*MSE*	*p*	*n2*
Written Annotations	6.55 (1,64)	200.02	<0.05	.093
Pictorial Annotations	12.16 (1,64)	371.28	<0.01	.160
Written and Pictorial Annotations	7.15 (1,64)	218.40	<0.05	.100

There was a significant main effect for pictorial and for written annotations, and a significant interaction effect between the two. Mean scores showed that the pictorial annotations group performed the best while the control group performed the poorest (Table 10).

Table 10. Mean Scores for the Four Treatment Groups on the Delayed Pictorial Vocabulary Recognition Posttest, Study 1

Groups	*N*	*M*	*SD*
Control	17	10.59	4.36
Pictorial Annotations	19	18.95	4.71
Written Annotations	18	17.50	5.59
Pictorial and Written Annotations	14	18.64	6.40

Post hoc comparisons (Tukey HSD) showed that all annotation groups outperformed the control group ($p<0.01$), but did not differ significantly from each other.

Students in the pictorial and written annotations group accessed the two annotation types with comparable frequency: 53% of the time with an average of 7.60 seconds per annotation for pictorial, and; 47% of the time with an average of 8.1 seconds per annotation for written annotations. The single annotation groups

viewed their respective annotations for equal amounts of time: 11.35 seconds for pictorial, and 11.51 seconds for written annotations.

In summary, all annotation groups performed significantly better than the control group on all tests. No other significant differences were found.

Immediate Written Vocabulary Production Posttest, Study 2

A 2 x 2 ANOVA was computed with the number of correct answers on the immediate written vocabulary production posttest as the dependent measure, and the presence or absence of pictorial and written annotations as the between subjects factor (Table 11).

Table 11. ANOVA for the Immediate Written Vocabulary Production Posttest, Study 2

Factors	*F*	*MSE*	*p*	*n2*
Written Annotations	93.6 (1,63)	1874.34	<0.001	.598
Pictorial Annotations	4.00 (1,63)	80.17	<0.05	.06
Written and Pictorial Annotations	5.07 (1,63)	101.49	<0.05	.074

There was a significant main effect for written and pictorial annotations and a significant interaction between the two types of annotations. The pictorial and written annotations and the written annotations groups performed the best while the control group performed the poorest (Table 12).

Table 12. Mean Scores of the Four Groups on the Immediate Written Vocabulary Production Posttest, Study 2

Groups	*N*	*M*	*SD*
Control	16	3.31	1.66
Pictorial Annotations	17	8.47	3.48
Written Annotations	18	16.33	5.29
Pictorial and Written Annotations	16	16.56	6.05

Post hoc comparisons (Tukey HSD) showed that the written annotations group did not differ significantly from those with access to both annotations, but that the written annotations and the pictorial and written annotations groups performed significantly better than did the pictorial annotations group, (p<0.001). All annotation groups performed significantly better than the control group.

Delayed Written Vocabulary Production Posttest, Study 2

A 2 x 2 ANOVA was computed with the number of correct answers on the delayed written vocabulary production posttest as the dependent measure and the presence or absence of pictorial and written annotations as the between subjects factor (Table 13).

Table 13. ANOVA for the Delayed Written Vocabulary Production Posttest, Study 2

Factors	*F*	*MSE*	*p*	*n2*
Written Annotations	40.42 (1,47)	367.4	<0.001	.462
Pictorial Annotations	0.096 (1,47)	0.872	<0.758	.002
Written and Pictorial Annotations	6.8 (1,47)	61.83	<0.050	.126

There was a significant main effect for written annotations and significant interaction effect between pictorial and written annotations. The written annotations group performed the best while the control group performed the poorest (Table 14).

Table 14. Mean Group Scores on the Delayed Written Vocabulary Production Posttest, Study 2

Groups	*N*	*M*	*SD*
Control	13	2.77	1.09
Pictorial Annotations	14	5.43	2.17
Written Annotations	13	10.31	4.03
Pictorial and Written Annotations	11	8.55	3.96

Post hoc comparisons (Tukey HSD) showed that the written annotations group did not differ significantly from those with access to both annotation types. The written annotations group and the pictorial and written annotations group had significantly higher scores than the control group ($p<0.001$). There was no significant difference between the pictorial annotations group and the written and pictorial annotations group. The difference between the pictorial annotations group and the control group was also not significant.

In terms of time on task, students in the pictorial and written annotations group did not access both annotation types with equal frequency: Pictorial annotations were accessed 37% of the time with an average of 7.01 seconds per annotation; written annotations were accessed 63% of the time with an average of 8.23 seconds per annotation. However, both annotation types were viewed for almost equal amounts of time by the single annotation groups: 10.98 seconds for pictorial and 11.21 seconds for written annotations.

In summary, the control group performed the poorest on both posttests. On the immediate written vocabulary production posttest, subjects who accessed both annotation types or written annotations alone outperformed those without access to written annotations. On the delayed test, the written annotations group retained more vocabulary than all other groups, while the pictorial annotations group did not differ significantly from the control group. Those who had access to written annotations alone or combined with pictorial annotations significantly outperformed those who did not have access to any written annotations.

DISCUSSION

Hypotheses 1 and 2 predicted that students with access to pictorial and written annotations during a L2 listening comprehension activity would recognize more written translations and pictorial representations of keywords on written vocabulary and pictorial vocabulary recognition posttests. These two hypotheses further predicted that students who accessed written annotations would outperform those without access to such annotations on the written vocabulary recognition posttest, while students who accessed pictorial annotations would outperform those without access to such annotations on the pictorial vocabulary recognition posttest.

The results of the immediate vocabulary recognition tests did not support these hypotheses because students recognized vocabulary equally well, regardless of test mode. Within recognition tests, there is an inherent ability to guess built into the testing format. Thus, previous exposure to the translation, either visually or verbally, makes selecting the correct response much easier than if one is asked to produce a response from memory (Cariana & Lee, 2001; Glover, 1989; McDaniel & Mason, 1985).

Hypothesis 3 predicted that students with access to pictorial and written annotations during a L2 listening comprehension activity would recall more vocabulary on a written vocabulary posttest than those without access to both annotation types, and also that students who accessed written annotations would outperform those without access to such annotations. Results of the immediate vocabulary production test show that the pictorial and written annotations group and the written annotations group recalled more vocabulary than did those without access to written annotations. This is in line with the third hypothesis and demonstrates that students learned more vocabulary when the testing mode employed matched the mode accessed, either alone or combined with an additional annotation mode.

With regards to all three hypotheses, the control group performed the poorest because the difficulty of the aural text prevented students from building contextual knowledge, thus lessening their ability to learn vocabulary incidentally (Hulstijn, 1992; Jones, 2003; Jones & Plass, 2002). On the other hand, vocabulary acquisition was consistently strong when students had access to pictorial and written annotations, thus supporting a multimedia effect proposed by Mayer (2001). The ability to look up words more than once in different modalities supported inferencing and verification strategies (Grace, 1998) and, reinforced learning (Chun & Plass, 1996), so that students were able to perform well on immediate tests regardless of testing mode. Additionally, students could establish direct connections between the L1 and L2 vocabulary and the corresponding images and thereby have two instead of just one retrieval route (Plass et al., 1998). However, with regard to the third hypothesis, students in the pictorial and written annotations group may have had too much information to look up and may have foregone examining both annotation types (Jones, 1995). Tracking logs showed that the pictorial and written annotations group did not examine the two types of annotations in a balanced manner, and this group subsequently performed poorer on the delayed written production test compared to the written annotations group. Though this group initially obtained a richer and redundant amount of information that was immediately helpful for producing written translations, with time, the retained information may have become "cluttered" and inhibited the students' ability to focus directly on the needed responses due to cognitive overload (Sweller, 1994).

Some researchers have argued that images carry a structural message that complements the language presented (Baggett, 1989; Kozma, 1991) and that the pictorial mode facilitates vocabulary learning (Kellogg & Howe, 1971; Oxford & Crookall, 1990; Underwood, 1989). This was the case in the study conducted by Jones and Plass (2002) in which students who accessed pictorial annotations alone or combined with written annotations outperformed those without access to any pictorial annotations on a written vocabulary recognition posttest. In the present study, students performed well no matter which annotation type was accessed. However, the pictorial annotations group could not produce vocabulary from memory as well as those groups that had access to written annotations, a result counter to findings that the pictorial mode of information increases the efficiency of learning (Kost, Foss & Lenzini, 1999; Oxford & Crookall, 1990; Terrell, 1986). Instead, images may have provided too much information (Sweller, 1994) rather than the more precise information provided by direct translations.

> There are more connections in the memory representation when the input is visual. "Brown leaf" presented verbally creates the instance of "leaf" connected with the concept "brown." But showing a picture of a brown leaf causes one to create the concept of leaf connected with concepts of brown, olive, rust, burgundy, etc., not to mention its shape, size, environment, etc. In the verbal presentation there is one sure connection: leaf with brown. (Baggett, 1989, p. 119)

The richness of images may have affected students' ability to accurately translate L2 words into L1, while written annotations provided precise definitions of the L2 words.

CONCLUSION

One limitation of this research is that it included two different studies with two different groups. Originally, these studies explored unrelated questions but once analyzed, the findings in terms of the influence of annotation types and testing modes on students' incidental vocabulary learning warranted a joint report. Further comparative analyses to determine the impact of pictorial and written annotations on pictorial and written testing modes was not possible since the two studies involved different subjects. An additional limitation is that both studies focused on a between-subjects design. A within-subjects examination may show how choice of annotation types affects students' performance on different vocabulary tests. Additional research should also consider the ordering of the tests. In Study 1, the pictorial recognition test was always administered before the written recognition test. Future examination using these two tests could be strengthened by counterbalancing their order.

These studies offer several implications for language teaching and multimedia design. They add to the growing body of literature on the beneficial effects of annotations on L2 vocabulary recognition and recall. The ability to review information more than once reinforces learning (Chun & Plass, 1996), and since students rely on different modalities to learn efficiently in different ways (Ehrman, Leaver, & Oxford, 2003; Oxford & Ehrman, 1995; Plass et al., 1998; Reinert, 1976), providing them with an opportunity to choose the mode of information they prefer may help them better learn the vocabulary presented. Multimedia environments that provide both pictorial and written modes of keyword information may be most effective for learning because the students can choose the mode that best suits their needs and learning preferences (Plass et al., 1998).

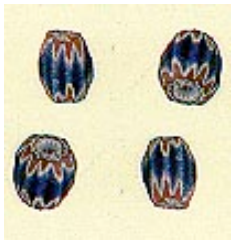
Future research is needed to examine issues related to the outcomes of these studies. For example, a more comparative approach to examining the connection between testing mode and annotation mode seems warranted. Further study is also needed to address the issues of cognitive load and the role it may play in long term memory for vocabulary when students access both pictorial and written annotations. Future study should also examine the recall of vocabulary in an aural mode rather than a written mode to determine how well aural L2 vocabulary comprehension is developed in an aural multimedia environment. And finally, to address vocabulary acquisition, researchers should examine students' ability to produce newly acquired vocabulary in a more challenging communicative context.

APPENDIX A
Text of Listening Comprehension Passage

LaSalle Meets the Quapaws (1682)

On fit traverser les canots sur la gauche à une pointe de sable. On se retranche le mieux que l'on peut avec des petits bois de tremble, qu'on coupa, dont on fit des palissades. La brume se cessa, et l'on vit un canot de Sauvages venir à nous …. Mais voyant qu'on ne leur en tirait point, ils s'en retournèrent chercher dire que c'étaient des gens de paix. Ils revinrent 6 sans armes avec le calumet de paix faisant signe qu'on vienne à leurs habitations. Ils présentèrent à fumer à M. de LaSalle et à tous ceux qui étaient autour de lui disant toujours qu'on s'embarque …. Le lendemain les guerriers et la jeunesse dansèrent le calumet. C'est de s'assembler tous sur la place. Les guerriers mettent leur présents sur des perches comme quand on veut faire sécher du linge. Ils apportent deux grands calumets enjolivés de plumes de toutes couleurs et pleins de cheveux de leurs ennemis. Ils mettent tout cela entre les mains de leurs chefs qui sont assis sur leurs culs et arrangés autour de la place. Ils ont tous des gourdes pleines de cailloux et des tambours qui sont des pots de terre. Ils commencent une chanson qu'ils accompagnent du bruit de leurs instruments. Les guerriers qui ont fait de belles actions vont danser et frapper un poteau avec leurs casse-tête et disent les belles actions qu'ils ont faites …. cependant les chefs fument, l'un après l'autre dans les calumets, et chacun le présentait à M. de LaSalle et à tous ceux dans la compagnie. Après, ils le prirent et le placèrent au milieu de la place, dansant tous autour de lui au son des instruments et chansons, chacun lui mettant sur le corps son présent qui étaient des peaux de boeufs qui ont de la laine comme nos moutons d'Europe. Si les Français ne l'avaient pas déchargé à mesure de ses peaux, ils l'auraient étouffé sous leurs présents. Il leur fit à son tour présents de haches, couteaux, et rassades.

APPENDIX B
Examples of 10 Images Used in the Pictorially-Based Treatments of Both Study 1 and Study 2

déchargé 	enjolivés 	étouffé 	sans armes 	bois 
le calumet 	la brume 	paix 	des rassades 	des cailloux 

APPENDIX C
Dependent Measures for Study 1 (Pictorial and Written Vocabulary Recognition Tests)

Name: _____ **Written Vocabulary Recognition Test**

In this activity, please select the English translation for the word given in French. If you do not know the correct response, leave it blank. You have 8 minutes to complete this exercise.

1. des rassades	6. la brume	11. casse-tête	16. enjolivés	21. des bois
a. poles	a. fog	a. hair	a. decorated	a. trees
b. cards	b. clouds	b. club	b. ugly	b. bodies
c. arrows	c. rain	c. arrow	c. plain	c. beads
d. beads	d. shells	d. pole	d. loud	d. chiefs
e. shells	e. water	e. pipe	e. noisy	e. pipes
f. nails	f. greeting	f. hatchet	f. rich	f. gourds
2. des canots	7. des perches	12. des cailloux	17. peaux de boeufs	22. culs
a. arrows	a. fish	a. poles	a. pears	a. poles
b. canoes	b. pipes	b. cards	b. skins	b. arrows
c. pipes	c. plazas	c. pebbles	c. sheep	c. hatchets
d. swords	d. presents	d. backsides	d. pigs	d. chiefs
e. canons	e. poles	e. shells	e. paddles	e. backsides
f. shells	f. cards	f. sheep	f. canoes	f. sheep
3. des haches	8. la laine	13. sans armes	18. le calumet	23. s'assembler
a. knives	a. pole	a. unarmed	a. pole	a. smoke
b. poles	b. card	b. angry	b. post	b. assemble
c. beads	c. arrow	c. tired	c. pipe	c. sit
d. helmets	d. bead	d. helpful	d. hatchet	d. dance
e. hatchets	e. shell	e. worried	e. knife	e. share
f. gourds	f. wool	f. artful	f. bead	f. smoke
4. sable	9. paix	14. les guerriers	19. un poteau	24. se retranche
a. sabers	a. peace	a. warriors	a. post	a. surround
b. sand	b. power	b. chiefs	b. card	b. retrain
c. swords	c. armed	c. pipes	c. bead	c. share
d. sheep	d. pipe	d. arrows	d. pot	d. provide
e. beads	e. canoe	e. canoes	e. drum	e. offer
f. knives	f. pole	f. turkeys	f. pipe	f. smoke
5. étouffé	10. les moutons	15. des gourdes	20. fument	25. déchargé
a. discarded	a. bullets	a. backsides	a. anger	a. weakened
b. suffocated	b. tables	b. drums	b. smear	b. offered
c. waved	c. arrows	c. gourds	c. shoot	c. smoked
d. helped	d. beef	d. grapes	d. smoke	d. removed
e. greeted	e. shells	e. pipes	e. trade	e. danced
f. offered	f. sheep	f. knives	f. give	f. played

Name: _____ **Pictorial Vocabulary Recognition Test Answer Sheet**

1. ____
2. ____
3. ____
4. ____
5. ____
6. ____
7. ____
8. ____
9. ____
10. ____
11. ____
12. ____
13. ____
14. ____
15. ____
16. ____
17. ____
18. ____
19. ____
20. ____
21. ____
22. ____
23. ____
24. ____
25. ____

Name: _____ **Pictorial Vocabulary Recognition Test**

In this activity, please select the image that represents the word given in French. Write the letter that represents your answer on the answer sheet provided. If you do not know a word at all, leave it blank. You have 8 minutes to complete this test.

[Sample of three questions.]

1. déchargé	A. 	B. 	C. 	D. 	E. 
2. étouffé	A. 	B. 	C. 	D. 	E. 
3. sans armes	A. 	B. 	C. 	D. 	E. 

APPENDIX D
Dependent Measures for Study 2 (Written Vocabulary Production Test)

Name: _____ **Written Vocabulary Production Test**

In this activity, please provide the English translation for the word given in French. If you do not know the response at all, leave it blank. You have 8 minutes to complete this test.

1. des rassades _____
2. des canots _____
3. des haches _____
4. sable _____
5. étouffé _____
6. la brume _____
7. des perches _____
8. la laine _____
9. paix _____
10. les moutons _____
11. casse-tête _____
12. des cailloux _____
13. sans armes _____
14. les guerriers _____
15. des gourdes _____
16. enjolivés _____
17. peaux de boeufs _____
18. le calumet _____
19. un poteau _____
20. fument _____
21. des bois _____
22. culs _____
23. s'assembler _____
24. se retranche _____
25. déchargé _____

ABOUT THE AUTHOR

Linda Jones is Assistant Professor of Instructional Technology and Director of the Language Learning Center at the University of Arkansas. Her principal interests include multimedia design theory and L2 learning, and instructing graduate students in the appropriate use of instructional system design strategies when developing L2 video and Web-based materials.

E-mail: lcjones@uark.edu

REFERENCES

Adobe. (1994). *Premiere* [Computer software]. Mountain View, CA: Adobe Systems.

Amon, E., Muyskens, J. A., & Omaggio Hadley, A. C. (2000). *Vis à Vis, 2nd Edition* [Computer software, CD-ROM]. Boston: McGraw-Hill Companies.

Baggett, P. (1989). Understanding visual and verbal messages. In H. Mandl & J. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 101-124). Amsterdam: Elsevier.

Baltova, I. (1999). The effect of subtitled and staged video input on the learning and retention of content and vocabulary in a second language. Unpublished doctoral dissertation. University of Toronto, Canada.

Brett, P. (1995). Multimedia for listening comprehension: The design of a multimedia-based resource for developing listening skills. *System, 23*(1), 77-85.

Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. *System, 25*(1), 39-53.

Buzhardt, G. A., & Hawthorne, M. (1993). *Rencontres sur le Mississippi, 1682-1763* [Encounters on the Mississippi, 1682-1763]. Jackson: University Press of Mississippi.

Cariana, R. B., & Lee, D. (2001). The effects of recognition and recall study tasks with feedback in a computer-based vocabulary lesson. *Educational Technology Research & Development 49*(3), 23-36.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*(2), 183-198.

Chun, D. M., & Plass, J. L. (1997). *Cyberbuch* [Computer Software, CD-ROM]. New York: St. Martin's Press.

Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition, 13*(4), 431-469.

Ehrman, M. E., Leaver, B. L., & Oxford, R. L. (2003). A brief overview of individual differences in second language learning. *System, 31*(3), 313-330.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*(3), 392-399.

Grace, C. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning-level CALL software. *The Modern Language Journal, 82*(4), 533-544.

Guillory, H. G. (1998). The effects of keyword captions to authentic French video on learner comprehension. *CALICO, 15*(1-3), 89-108.

Herron, C. (1994). An investigation of the effectiveness of using an advance organizer to introduce video in the foreign language classroom. *The Modern Language Journal, 78*(2), 190-198.

Herron, C., York, H., Cole, S., & Linden, P. (1998). A comparison study of student retention of foreign language video: Declarative versus interrogative advance organizers. *The Modern Language Journal, 82*(2), 237-247.

Hoven, D. (1999). A model for listening and viewing comprehension in multimedia environments. *Language Learning & Technology, 3*(1), 88-103. Retrieved March 30, 2004, from http://llt.msu.edu/vol3num1/hoven/

Hulstijn, J. H. (1989). Implicit and incidental second language learning: Experiments in the processing of natural and partly artificial input. In H. W. Dechert & M. Raupach (Eds.), *Interlingual processes* (pp. 49-73). Tübingen, Germany: Gunter Narr.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London: Macmillan.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*(3), 327-339.

Jonassen, D., & Tessmer, M. (1996). An outcome-based taxonomy for instructional systems design, evaluation and research. *Training Research Journal, 2,* 11-46.

Jones, F. (1995). Learning an alien lexicon: A teach-yourself case study. *Second Language Research, 11*(2), 95-111.

Jones, L. (2003). Supporting listening comprehension and vocabulary acquisition with multimedia annotations: The students' voice. *CALICO, 21*(1), 41-65.

Jones, L., & Plass, J. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal, 86*(4), 546-561.

Kellogg, G., & Howe, M. J. (1971). Using words and pictures in foreign language learning. *Alberta Journal of Educational Research, 17*(2), 87-94.

Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *The Modern Language Journal, 78*(3), 285-299.

Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals, 32*(1), 89-113.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61*(2), 179-211.

Larson, J., & Bush, C. (1992). *Montevidisco* [Computer software]. Provo, UT: Brigham Young University.

Lynch, T. (1998). Theoretical perspectives on listening. *Annual Review of Applied Linguistics, 18,* 3-19.

Macromedia. (1997). *Authorware 4.0* [Computer software]. San Francisco, CA: Macromedia, Inc.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1-19.

Mayer, R. E. (2001). *Multimedia learning.* Cambridge, England: Cambridge University Press.

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: An example of the two-way street between cognition and instruction. *New Directions for Teaching and Learning, 89,* 55-71.

McDaniel, M. A., & Mason, M. E. J. (1985). Altering memory representations through retrieval. Journal of experimental psychology. *Learning, Memory and Cognition, 11,* 371-385.

Otto, S., & Pusack, J. (1992). *Listening Tool* [Computer software]. Iowa City, IA: PICS.

Oxford, R., & Crookall, D. (1990). Vocabulary learning: A critical analysis of techniques. *TESL Canada Journal, 7*(2), 9-30.

Oxford, R. L., & Ehrman, M. E. (1995). Adults' language learning strategies in an intensive foreign language program in the United States. *System, 23*(3), 359-386.

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford, England: Oxford University Press.

Pica, T., Doughty, C., & Young, R. (1986). Making input comprehensible: Do interactional modifications help? *International Review of Applied Linguistics, 72,* 1-25.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second language multimedia learning environment. *Journal of Educational Psychology, 90*(1), 25-36.

Reinert, H. (1976). One picture is worth a thousand words? Not necessarily. *The Modern Language Journal, 60*(4), 160-168.

Sabo, G., Restropo, L. F., & Jones, L. (2000). *First encounters: Native Americans and Europeans in the Mississippi Valley* [Computer software: CD-ROM]. Fayetteville, AR: Arkansas Archeological Society.

Salaberry, M. R. (2001). The use of technology for second language learning and teaching: A retrospective. *The Modern Language Journal, 85*(1), 39-56.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4,* 295-312.

Terrell, T. (1986). Acquisition in the natural approach: The binding/access framework. *The Modern Language Journal, 70*(3), 213-227.

Underwood, J. (1989). HyperCard and interactive video. *CALICO, 6*(3), 7-20.

University of Texas (2001). *Français Interactif* [Web Site]. http://www.lamc.utexas.edu/fi

University of Texas (2004). *Français Interactif* [Web Site]. http://www.laits.utexas.edu/fi

Wittrock, M. C. (1989). Educational psychology and the future of research in learning, instruction and teaching. In M.C. Wittrock & F. Farley (Eds.), *The future of educational psychology* (pp. 75-89). Hillsdale, NJ: Erlbaum.

Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO, 20*(1), 33-58.